



Original Article

Generosity is a sign of trustworthiness—the punishment of selfishness is not

Wojtek Przepiorka^{a,b,*}, Ulf Liebe^c^a Department of Sociology, Utrecht University, Padualaan 14, Utrecht 3584 CH, The Netherlands^b Nuffield College, New Road, Oxford OX1 1NF, United Kingdom^c Institute of Sociology, University of Bern, Fabrikstrasse 8, CH-3012 Bern, Switzerland

ARTICLE INFO

Article history:

Initial receipt 27 July 2015

Final revision received 28 December 2015

Keywords:

Peer-punishment

Third-party punishment

Signaling theory

Generosity

Trust

Trustworthiness

ABSTRACT

Peer-punishment is an important determinant of cooperation in human groups. It has been suggested that, at the proximate level of analysis, punitive preferences can explain why humans incur costs to punish their deviant peers. How punitive preferences could have evolved in humans is still not entirely understood. A possible explanation at the ultimate level of analysis comes from signaling theory. It has been argued that the punishment of defectors can be a type-separating signal of the punisher's cooperative intent. As a result, punishers are selected more often as interaction partners in social exchange and are partly compensated for the costs they incur when punishing defectors. A similar argument has been made with regard to acts of generosity. In a laboratory experiment, we investigate whether the punishment of a selfish division of money in a dictator game is a sign of trustworthiness and whether punishers are more trustworthy interaction partners in a trust game than non-punishers. We distinguish between second-party and third-party punishment and compare punitive acts with acts of generosity as signs of trustworthiness. We find that punishers are not more trustworthy than non-punishers and that punishers are not trusted more than non-punishers, both in the second-party and in the third-party punishment condition. To the contrary, second-party punishers are trusted less than their non-punishing counterparts. However, participants who choose a generous division of money are more trustworthy and are trusted more than participants who choose a selfish division or participants about whom no information is available. Our results suggest that, unlike for punitive acts, the signaling benefits of generosity are to be gained in social exchange.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In the last three decades, the literature on peer-punishment as a mechanism to sustain cooperation in humans has thrived (Oliver, 1980; Axelrod, 1986; Boyd & Richerson, 1992). In situations in which individuals have an incentive to free ride on others' cooperative efforts, the presence of group members who punish free riders at an immediate cost to themselves can promote and maintain cooperation in the group (Ostrom, Walker, & Gardner, 1992; Fehr & Gächter, 2002; Güerke et al., 2006). If the benefits of a cooperative environment outweigh the costs of maintaining a credible punishment threat, then peer-punishment is both rational and fitness enhancing and thus can be explained both at the proximate and ultimate level of analysis, respectively (Gächter et al., 2008; Raihani & Bshary, 2011; Przepiorka & Diekmann, 2013; Roberts, 2013). However, costly peer-sanctioning has been observed in one-time-only encounters between unrelated individuals both in the lab (Fehr & Fischbacher, 2004; Diekmann & Przepiorka, 2015) and

in the field (Henrich et al., 2006; Balafoutas, Nikiforakis, & Rockenbach, 2014; Diekmann, Jann, Przepiorka, & Wehrli, 2014). In these situations, the benefits of peer-punishment are unlikely to outweigh the costs and, therefore, peer-punishment cannot be readily explained from within the rational choice and the individual-selectionist framework, respectively (Hamilton, 1963; Trivers, 1971; Becker, 1976). Punitive preferences have been proposed as a proximate explanation for why humans sanction their peers even in situations in which they incur a net loss (Gintis, 2000; Fehr, Fischbacher, & Gächter, 2002), and there is an ongoing debate about the function such punitive preferences evolved to fulfill (Sigmund, 2007; Dreber, Rand, Fudenberg, & Nowak, 2008; Baumard, 2010; Boyd, Gintis, & Bowles, 2010; Raihani & Bshary, 2011; West, El Mouden, & Gardner, 2011; Barclay, 2012; Guala, 2012; Krasnow, Cosmides, Pedersen, & Tooby, 2012).

One explanation for the evolution of punitive preferences which has received little attention comes from signaling theory (Spence, 1974; Zahavi, 1975, 1977; Bliege Bird & Smith, 2005; Gambetta, 2009). It has been argued that pro-social acts can function as a type-separating signal of an individual's unobservable quality, if this quality is causally related to the individual's ability to cooperate (Zahavi, 1995; Gintis, Smith, & Bowles, 2001). This argument has received empirical support. It has

* Corresponding author. Department of Sociology, Utrecht University, Padualaan 14, Utrecht 3584 CH, The Netherlands. Tel.: +31 30 253 3467.

E-mail address: w.przepiorka@uu.nl (W. Przepiorka).

been shown that generosity can be a type-separating signal of an individual's trustworthiness (Barclay, 2004; Fehrler & Przepiorka, 2013; Gambetta & Przepiorka, 2014; Bliche Bird & Power, 2015) and that the signaling benefits of altruistic acts which accrue in social exchange can ease the conditions under which other-regarding preferences can evolve (Delton, Krasnow, Cosmides, & Tooby, 2011; Fehrler & Przepiorka, 2013). Since peer-punishment is often conceived as pro-social or even altruistic (Fehr & Gächter, 2002; Boyd, Gintis, Bowles, & Richerson, 2003; de Quervain et al., 2004; Fowler, 2005), it has been argued that peer-punishment too could work as a type-separating signal of the punisher's cooperative intent (Gintis et al., 2001). There is further theoretical support for this argument.

Many evolutionary models which show that punishment can promote the evolution of cooperation devise conditions under which cooperators who also punish defectors constitute an evolutionary stable strategy (Boyd et al., 2003; Hauert, Traulsen, Brandt, Nowak, & Sigmund, 2007; Helbing, Szolnoki, Perc, & Szabó, 2010; dos Santos, Rankin, & Wedekind, 2011). Under these conditions, punishment and (first-order) cooperation will be correlated and thus the former will be a reliable sign of the latter. Although there is empirical evidence in support of a signaling account of peer-punishment (Barclay, 2006; Kurzban, DeScioli, & O'Brien, 2007; Nelissen, 2008; Simpson, Harrell, & Willer, 2013), there is also evidence opposing or not supporting it (Rockenbach & Milinski, 2011; Pedersen, Kurzban, & McCullough, 2013; Balafoutas et al., 2014; FeldmanHall, Sokol-Hessner, Bavel, & Phelps, 2014; Gordon, Madden, & Lea, 2014). Based on a comprehensive review of this literature, a more elaborate argument recently emerged which tries to pin down the conditions under which we can expect peer-punishment to be a sign of a punisher's cooperative intent (Raihani & Bshary, 2015a).

The conceptual framework put forward by Raihani and Bshary (2015a) is informed by the growing literature investigating the proximate mechanisms behind individuals' punitive acts (Xiao & Houser, 2005; Jordan, McAuliffe, & Rand, 2015; Bone & Raihani, 2015). Peer-punishment can be triggered by different motives across different contexts (Leibbrandt & López-Pérez, 2012). Thus, punitive acts may be ambiguous in the information they convey about punishers' underlying motivations (Brañas-Garza, Espín, Exadaktylos, & Herrmann, 2014; Raihani & Bshary, 2015a). In particular, it has been argued theoretically and shown empirically that different motives might trigger peer-punishment in so-called second-party and in third-party punishment situations (Fehr & Fischbacher, 2004; Henrich et al., 2006; Carpenter & Matthews, 2009; Marlowe et al., 2011; Rockenbach & Milinski, 2011; Leibbrandt & López-Pérez, 2012; FeldmanHall et al., 2014; Gummerum & Chu, 2014; Harris, Herrmann, Kontoleon, & Newton, 2015). Punishing a deviant peer on one's own behalf (second-party punishment) is more likely to be motivated by vengefulness and thus more likely to be perceived as such by an observer (Marlowe et al., 2011; Rockenbach & Milinski, 2011). Punishing a deviant peer on the part of another "victim" (third-party punishment) is more likely to be motivated by the normative desire to establish justice and more likely to be perceived as such by an observer (Willer, Kuwabara, & Macy, 2009; Simpson et al., 2013; FeldmanHall et al., 2014).

1.1. Research question and hypotheses

Here we address the question whether peer-punishment can function as a type-separating signal of a punisher's cooperative intent. There are two necessary conditions for a signaling account of peer-punishment to be plausible. First, punitive preferences and cooperative intent must be positively related. Second, observers must infer cooperative intent from punitive acts. We conduct a laboratory experiment with economic games to test whether these two conditions are met. In our experiment, we measure subjects' punitive preferences in terms of their decisions to punish another subject for a selfish (i.e. self-regarding) division of money in a binary dictator game, we measure subjects' cooperative intent in terms of their trustworthiness as second movers in a trust game, and

we measure whether subjects infer trustworthiness from punitive acts by these subjects' trust as first movers in the trust game. Trustworthiness is a concept widely used in the social sciences and stands for the cooperative intent of the second-moving party in social exchange (Coleman, 1990; Hardin, 2002; Gambetta & Hamill, 2005; Fehr, 2009). Our first two hypotheses can be stated as follows:

- H1.** *Actors who punish selfish behavior are more trustworthy than actors who do not punish selfish behavior.*
- H2.** *Actors who punish selfish behavior are trusted more than actors who do not punish selfish behavior.*

Based on the literature cited above, we expect that third-party punishment is a better sign of trustworthiness than second-party punishment, because we expect that a sense of justice sustains trustworthiness better than vengefulness does (Marlowe et al., 2011; Raihani & Bshary, 2015a). However, these two motives cannot be readily separated. For example, it cannot be ruled out *a priori* that a sense of justice will sometimes trump vengefulness in motivating second-party punishment. In the context of kinship relations and close friendships, vengefulness may also trump a sense of justice in motivating third-party punishment. We reduce the likelihood of vengefulness to motivate third-party punishment to a minimum by design. In our laboratory experiment, subjects interact with each other from behind their computer screens while sitting in isolated cubicles; the anonymous environment reduces the ability of third-party observers to empathize with recipients of a selfish division in the dictator game. Accordingly, our next two hypotheses are as follows:

- H3.** *The positive relation hypothesized under H1 is stronger with regard to third-party punishment than with regard to second-party punishment.*
- H4.** *The positive relation hypothesized under H2 is stronger with regard to third-party punishment than with regard to second-party punishment.*

Finally, we compare the information punitive acts convey with the well-established finding that generosity is positively related with trustworthiness and observers infer trustworthiness from acts of generosity. We call the more equal division of money in our binary dictator game "generous," although it need not be motivated by generosity alone, but could also be motivated by a sense of fairness or the adherence to a social norm for sharing; what matters is that all these motives too can sustain trustworthiness (Gambetta & Przepiorka, 2014). Hence, our last two hypotheses can be stated as follows:

- H5.** *Actors who are generous are more trustworthy than actors who are selfish.*
- H6.** *Actors who are generous are trusted more than actors who are selfish.*

To our knowledge, this is the first experimental study to directly compare punitive acts and acts of generosity as signs of trustworthiness. Given that both punitive acts and acts of generosity are important elements of human sociality, their relative importance as signs of trustworthiness will emerge from the direct comparison. At the time we conducted our experiment, we did not have any expectations as to whether generosity or punishment would prove to be the *better* sign of trustworthiness.

2. Materials and methods

2.1. Experimental games

We use the binary dictator game with second-party punishment (DG2P) and third-party punishment (DG3P) to measure subjects'

Download English Version:

<https://daneshyari.com/en/article/943118>

Download Persian Version:

<https://daneshyari.com/article/943118>

[Daneshyari.com](https://daneshyari.com)